

Video-based Side-view Face Recognition for Home Safety

Pinar Santemiz

Luuk J. Spreeuwers

Raymond N.J. Veldhuis

University of Twente

Signals and Systems Group, Department of Electrical Engineering

Drienerlolaan 5 P.O.Box 217 7500AE

Enschede, The Netherlands

p.santemiz@utwente.nl L.J.Spreeuwers@utwente.nl R.N.J.Veldhuis@utwente.nl

Abstract

In this paper, we introduce a registration method for side-view face recognition that is suitable for home safety applications. We use cameras attached at door posts, and recognize people as they pass through doors to estimate their location in the house. First, we present a new database that is collected using this setup, where we use side cameras and ambient light. We recorded videos of 14 people that pass through doors in 18 different paths. Next, we propose our recognition method where we automatically find the profile to register the face images. By applying hierarchical clustering we detect the frames that include falsely detected profiles and pose variations, and automatically remove them from the video sequence to improve our results. After registering, we find the nose tip, apply recognition based on profiles, and analyze our results.

1 Introduction

Face recognition is a biometric method with many applications for its nature of being non-intrusive, natural, and passive. Especially, in applications dealing with identifying people from videos, face recognition is the primary biometrics. However, it is a challenging task to recognize faces in real-life scenarios due to occlusion, expression, or pose variations.

Home safety applications are one of the possible implementation areas for face recognition. The accidents and injuries that happen in the home environment are mostly caused by overlooked risks, busy schedule of the parents, or external threats. Therefore, face recognition can be used to increase the situational awareness, to prevent the factors that may cause further accidents, or to detect an emergency in time.

In this paper we introduce a novel method for landmark detection for side-view face recognition to be used in house safety applications. Our aim is to identify people as they walk through doors, and estimate their location in a house. In our system, we will use video recordings from cameras that are attached to door posts. Due to the location of the cameras, the range of sight will be limited and people are only recorded as they walk through doors. Consequently, the privacy of the people will be preserved.

Side-view face recognition is a challenging problem due to the complex structure of human face. A literature survey on face recognition under pose variations can be found in [21]. The first attempts to compare side-view face images were based on comparing profile curves, where fiducial points or features describing the profile were used for recognition. One such method is proposed by Gao and Leung [7], where they match profile line segments, and apply Hausdorff distance to measure similarity. They achieve 96.7% recognition accuracy on the Bern Database [6], which contains side-view face images and silhouettes of 30 people. Bhanu and Zhou [4] propose a curvature-based matching approach, where they use the curvature values to find nasion and throat

point, and then compare the curvature values in-between using Dynamic Time Warping (DTW). They achieve a recognition accuracy of 90.00% on Bern database. In a later work [22], they propose a method to construct a high resolution face profile image from low resolution videos. They use an elastic registration algorithm for alignment of profiles, and apply recognition using DTW. They experiment on 28 video sequences of 14 people walking with a right angle to the camera, and recognize more than 70% of the people correctly.

In applications where identification of people from videos is aimed, the system should be able to handle pose variations. Therefore, in many approaches where videos are involved, people make use of the texture information in addition to profile curves. Tsalakanidou *et al.* [19] present a face recognition technique based on depth and color eigenfaces, where they use the depth map for exploiting the 3D information. They experiment on XM2VTS database using 40 subjects, and recognize 87.5% of them correctly. Gross *et al.* [10] investigate recognition of human faces in a meeting room, where they propose a method called Dynamic Space Warping (DSW). Here, they apply Principal Component Analysis (PCA) [20] to vectors of sub-images from a given face, and compare these sequences using dynamic programming. They evaluate their algorithm using recordings of six meetings with six people, and achieve an accuracy of 89.4% on images without occlusion.

Instead of handling the pose variation at feature level, in some applications the images are warped or synthesized using 2D or 3D-aided systems, so that images contain the same pose as the image that is compared to. An early approach was proposed by Beymer and Poggio [3], where they generate 14 virtual views of a given face, and use them together with the original example for enrollment. They compute the correlation between images using optical flow and template matching, and achieve a recognition accuracy of 70.20% in a database with 62 people using a cross-validation methodology. Blanz and Vetter [5] estimate 3D shape and texture of faces from single 2D images using a statistical, morphable 3D model. The results on CMU-PIE and FERET show that, the algorithm achieves 95.00% and 95.90% correct identifications, respectively. Kakadiaris *et al.* [14] presents a side-view face recognition system, where they make use of 3D face models for enrollment, and extract profiles under different poses. For recognition, they extract profiles from given images and use Vector Distance Function (VDF) to match the profiles to the gallery profiles. Their system achieves a 60.00% recognition accuracy on the database UHDB1.

In this study, we introduce our registration method for side-view face recognition. We will first present our new database that includes videos of 14 people passing through doors in Section 2. We recorded these videos in ambient light using cameras attached to door post. Then we will propose our registration method which relies on profile lines. The details of our registration method are given in Section 3. After finding the registration parameters, we compute the median profile and find the tip of the nose. We test our system against our new database and analyze our results in Section 4. Finally, we will give our conclusion in Section 5, and discuss our future work.

2 Database

In our study, we aim to recognize people as they pass through doors. We assume that the system will be used in home environment, in ambient light, and the person may approach the door from any direction. Therefore, our system should be robust to poor illumination conditions and pose variation. Moreover, the camera needs to have a limited view angle, so that it only records the person's face, and not invade the privacy of the household.

There are a number of databases that contain face images with variant poses including side-view face images [8]. Most of them are collected in a controlled environment with uniform background, artificial illumination changes, or restricted pose variations.

These databases mostly contain still images. Even though there are some databases, that contain videos of people in less controlled settings, they either contain small pose variations, or an unrealistic scenario. Therefore, we collected a database that contains videos of people passing through a door.

We attached a camera at the doorpost which has a limited view angle and only records the person's face from side-view. We used ambient light in our recordings, and used a resolution of 1024×768 pixels. We asked the people to approach the door from left, from right, or from opposing direction, and after passing through the door to continue either straight forward or to turn left or right. So, there are totally 18 recordings for each person, and 14 people in our database. The duration of each recording is 45 frames, which corresponds to one and a half seconds.

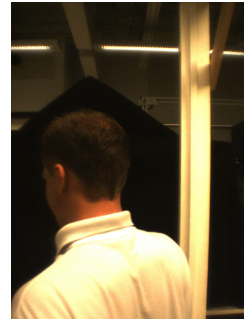
The duration for passing through the door may change from half a second to one and a half seconds depending on how the person approaches the door. Within this time, the face is only visible in five to 20 frames. As seen in the examples in Figure 1 there is a huge variation in head poses. Due to the position of the camera, in some videos the face is not visible. Therefore, we eliminated 40 recordings after visual inspection, and obtained a database with 216 recordings.



(a)



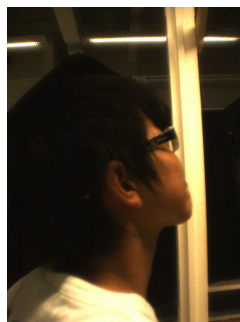
(b)



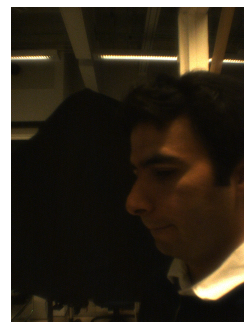
(c)



(d)



(e)



(f)

Figure 1: Examples of the recordings from our database.

3 Registration

Our registration method relies on matching profiles. Therefore we first find masks that only contain the face using skin color information and background subtraction. Then, we extract profiles from each frame of the recordings. These profiles may contain local errors due to poor illumination conditions, shadows, or large pose variation. Therefore, we first align the profiles in an image sequence to the profile of the central frame. Then we use hierarchical clustering to eliminate the most erroneous profiles, and compute the median profile for each recording. Finally, using the curvature along this median profile, we find the tip of the nose. The details to our face segmentation method is presented in Section 3.1, our profile extraction method is given in Section 3.2, and our method for finding the tip of the nose is in Section 3.3.

3.1 Face segmentation

In order to obtain a mask that only contains the face, we first apply background subtraction, and then combine this with skin color information. For background subtraction, we use adaptive Gaussian mixture models [23]. Due to poor illumination conditions, this image contains a lot of noise. Moreover, it contains more information than just the face. Therefore we detect skin color using HSV color space, where we threshold the hue values. Then we combine the background mask with skin color mask, and remove the noise. Our example to our face segmentation algorithm is shown in Figure 2.

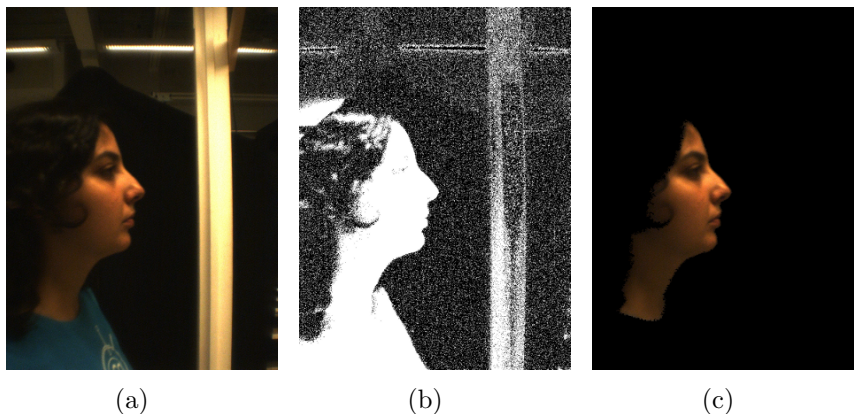


Figure 2: Segmentation of the face. (a) Original image. (b) Mask obtained after background subtraction. (c) Segmented face image.

3.2 Extracting Profile

After segmenting the face, we first find the direction in which the person is walking. In order to do that, we compute the center of the face in each frame, and from their differences we find the direction. Using the direction, we decide which half of the face contains the profile. Then, we extract the outer edge of the segmented face and remove the edge pixels that are located in the back of the head. An example is shown in Figure 3.

Due to poor illumination conditions, or shadows some local errors may occur in some profiles. Also, in the first and last few images some parts of the face might be missing. Moreover, when a large pose variation occurs within a recording, some profiles may be missing information of the face shape. In order to eliminate these profiles, we

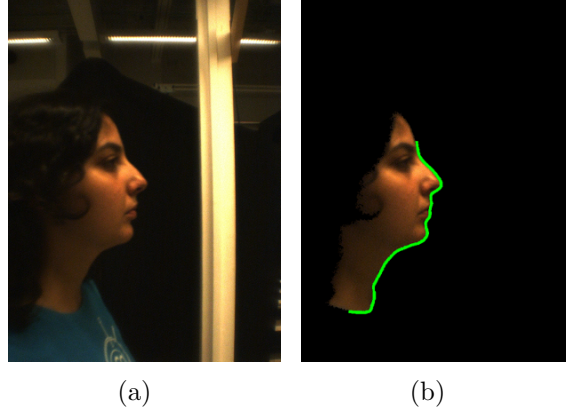


Figure 3: Extraction of the profile. (a) Original image. (b) Profile of the face.

first smooth the profiles and align each profile to the profile of the central frame using template matching. We assume that within a single recording, the face may have a tilt between -10 to 10 degrees. Also, the scale of the face might change due to the distance between the person and the camera. We assumed that the change of the scaling factor is between 0.8 and 1.2 . We match the profiles to the profile of the central frame using Hausdorff distance [11], and find the transformation parameters for each profile. Using these parameters, we align the profiles. Next, we compute the distances between each profile using Hausdorff distance, and apply hierarchical clustering to eliminate erroneous profiles. Finally, we compute the median of the remaining profiles to find the median profile. Figure 4 illustrates an example of the computation of the median profile.

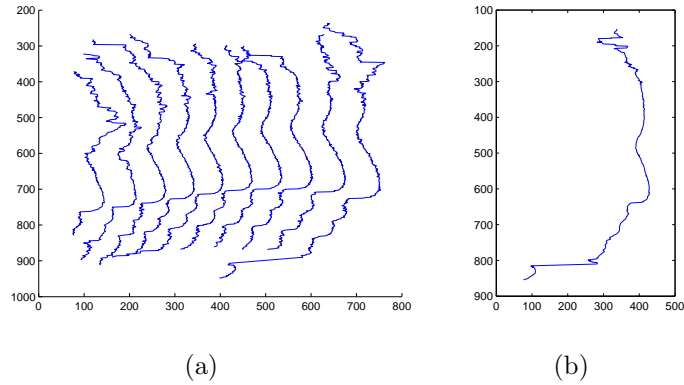


Figure 4: Finding the median profile. (a) Profiles obtained from an image sequence. (b) Median profile.

3.3 Finding the Nose Tip

In our experiments we assume that the tip of the nose is located near the center of the face, and has the maximum curvature. But in most of our profiles, neck is also included to the image. Therefore, we first remove the neck, and than search for the maximum curvature around the center of the face.

In order to find the neck, we first find the principle component of the edge pixels, and rotate the face according to this principle component. Then, we assume that between the top of the head and the eye edge pixels monotonically increase, and between the chin and the throat they monotonically decrease. According to this assumption, we determine the face. After finding the face, we search for the maximum curvature located around the center of the face. In Figure 5 an example of this method is illustrated.

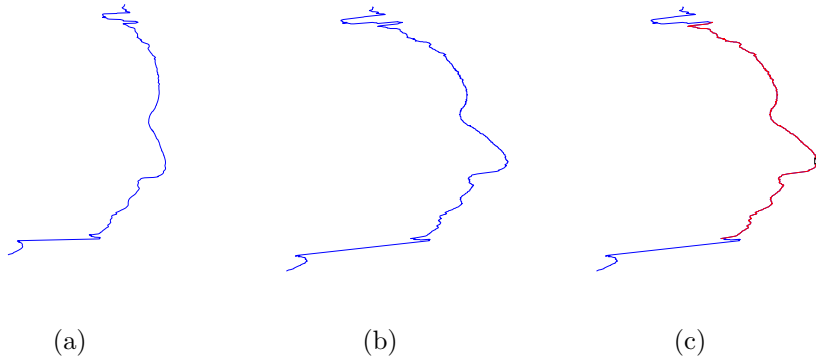


Figure 5: Obtaining the area where the face is located. (a) Median profile. (b) Median profile after rotating according to the principle component. (c) Location of the face and the tip of the nose.

In order to see the performance, we manually labeled the nose tips. In our database, the average length of the face is 500 pixels. Therefore, we assumed that if the nose tip is found within 7 per cent of the face size, our algorithm for finding the nose would be correct. According to this assumption, we were able to determine the nose-tip correctly in 207 recordings out of 216 recordings and achieved an accuracy of 95.83 per cent.

4 Experimental Results

In our experiments, we divided our set into two subsets: an enrollment set with one recording for each person, and a test set with a total of 202 recordings. In the videos belonging to the enrollment set the person is approaching the door from opposing direction and walks straight. Therefore, in these videos the pose variation is minimal. For recognition, we align each profile in the test set to each profile in the enrollment set using a modified template matching method. Here, we first translate the test profile so that the nose tip coordinates of the test profile coincides with the nose tip coordinate of the enrollment profile. Next, we rotate and scale the test profile for several parameters, and choose the parameters that gives the smallest Hausdorff distance as transformation parameters. A result of our modified template matching algorithm is shown in Figure 6.

When visually inspected, the alignment method successfully aligns the profiles belonging to the same person. Since the method does not rely on finding the landmarks, it is more robust to pose variations. However, when we use these distances for recognition, we achieve a very low accuracy of 22.78 per cent. This is because using the profiles alone is not reliable enough for recognizing the faces.

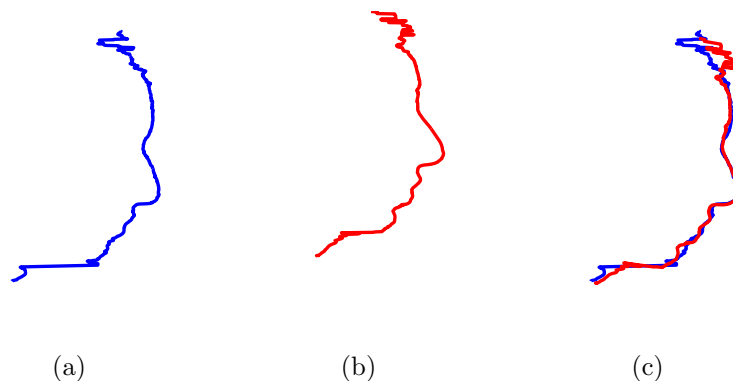


Figure 6: Modified Hausdorff distance method. (a) Enrollment profile. (b) Test profile. (c) Enrollment profile and aligned test profile.

5 Conclusion and Future Work

In this work we investigate a registration method for side-view face recognition to be used in house safety applications, where we aim to identify people as they walk through doors. Our method relies on profiles along the side-view face images, and only one landmark point on this profile. We first apply a combination of background subtraction and skin color detection to obtain a segmented face image for each frame. From these images we extract the profile lines, and after matching each profile to the profile of the central frame we apply hierarchical clustering to remove erroneous profiles. From the profiles of an image sequence, we compute the median profile, and using the curvatures along this median profile we find the tip of the nose.

When we compare profiles from different videos, we again apply template matching, but we use the tip of the nose to find the translation parameters directly. When visually inspected we see that our alignment algorithm is promising. However, comparing the aligned profiles for recognition results in poor accuracy. We conclude that shape information should be combined with the texture information. In the future we are planning to use Local Binary Patterns to extract the texture information and use it for recognition.

6 Acknowledgement

This work is supported by GUARANTEE (ITEA 2) 08018 project.

References

- [1] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE Transactions on PAMI*, vol.28, pp.2037–2041, 2006.
- [2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection,” *IEEE Transactions on PAMI*, vol.19, pp.711–720, 1997.
- [3] D. Beymer, and T. Poggio, “Face recognition from one example view,” *Computer Vision, IEEE Int. Conf. on*, pp.500–507, 1995.

- [4] B. Bhanu, and X. Zhou, "Face Recognition from Face Profile Using Dynamic Time Warping," *Int. Conf. on Pattern Recognition (ICPR)*, vol.4, pp.499–502, 2004.
- [5] V. Blanz, and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on PAMI*, vol.25, pp.1063–1074, 2003.
- [6] <ftp://ftp.iam.unibe.ch/pub/Images/FaceImages> .
- [7] Y. Gao, and M. Leung, "Line segment Hausdorff distance on face matching," *Pattern Recognition*, vol.35, pp.361–371, 2002.
- [8] R. Gross, "Face Databases," *Handbook of Face Recognition*, pp.301–327, 2005.
- [9] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Computing*, vol.28, pp.807–813, 2010.
- [10] R. Gross, J. Yang, and A. Waibel, "Face Recognition in a Meeting Room," *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp.294, 2000.
- [11] M.-P. Dubuisson and A.K. Jain, "A Modified Hausdorff Distance for Object Matching", *IEEE Int. Conf. on Computer Vision and Image Processing*, vol.1, pp.566–568, 1994.
- [12] <http://cbl.uh.edu/URxD/datasets/> .
- [13] <http://www.tele.ucl.ac.be/PROJECTS/M2VTS/m2fdb.html> .
- [14] I.A. Kakadiaris, H. Abdelmunim, W. Yang, and T. Theoharis, "Profile-based face recognition," *IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG '08)*, pp.1–8, 2008.
- [15] K. Messer, J. Matas, J. Kittler, J. Luetin, and G. Maitre, "XM2VTSDB: The Extended M2VTS Database," *Int. Conf. on Audio and Video-based Biometric Person Authentication*, pp.72–77, 1999.
- [16] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-Based Database for Facial Expression Analysis," *IEEE Int. Conf. on Multimedia and Expo*, pp.317–321, 2005.
- [17] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, and L. Akarun, "Bosphorus Database for 3D Face Analysis," *Biometrics and Identity Management* , vol.5372, pp.47–56, 2008.
- [18] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression Database," *IEEE Transactions on PAMI*, vol.25, pp.1615–1618, 2003.
- [19] F. Tsalakanidou, "Use of depth and colour eigenfaces for face recognition," *Pattern Recognition Letters*, vol.24, pp.1427–1435, 2003.
- [20] M. Turk, and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol.3, pp.71–86, 1991.
- [21] X. Zhang, and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol.42, pp.2876–2896, 2009.
- [22] X. Zhou, and B. Bhanu, "Human Recognition Based on Face Profiles in Video," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, pp.15, 2005.
- [23] Z.Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction", *Int. Conf. on Pattern Recognition*, vol.2, pp.28–31, 2004.